PROCEEDINGS OF THE ROYAL SOCIETY B | BIOLOGICAL SCIENCES

Linking the antigen archive structure to pathogen fitness in African trypanosomes

Erida Gjini, Daniel T. Haydon, J. D. Barry and Christina A. Cobbold

Supplementary data	"Data Supplement" http://rspb.royalsocietypublishing.org/content/suppl/2012/12/28/rspb.2012.2129.DC1.html
References	**This article cites 59 articles, 15 of which can be accessed free** http://rspb.royalsocietypublishing.org/content/280/1753/20122129.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections ecology (1234 articles) evolution (1363 articles) health and disease and epidemiology (193 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Linking the antigen archive structure to pathogen fitness in African trypanosomes

Erida Gjini[1,3,†], Daniel T. Haydon[2,3,4], J. D. Barry[4] and Christina A. Cobbold[1,3]

[1]School of Mathematics and Statistics, College of Science and Engineering, [2]Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, [3]The Boyd Orr Centre for Population and Ecosystem Health, and [4]Wellcome Trust Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK

Systems that generate antigenic variation enable pathogens to evade host immune responses and are intricately interwoven with major pathogen traits, such as host choice, growth, virulence and transmission. Although much is understood about antigen switching at the molecular level, little is known about the cross-scale links between these molecular processes and the larger-scale within and between host population dynamics that they must ultimately drive. Inspired by the antigenic variation system of African trypanosomes, we apply modelling approaches to our expanding understanding of the organization and expression of antigen repertoires, and explore links across these scales. We predict how pathogen population processes are determined by underlying molecular genetics and infer resulting selective pressures on important emergent repertoire traits.

1. Introduction

Understanding the population dynamics of pathogens is essential for controlling infectious disease. One key factor in infection dynamics is antigenic variation, an immune evasion strategy shared by bacteria, viruses and protozoa [1–5], posing a major impediment to infection control within [6,7] and between hosts [8]. Among protozoan parasites the most prominent examples are *Plasmodium falciparum*, the malaria parasite, and the African trypanosome, causative agent of human sleeping sickness and livestock disease in Africa. As a result of antigenic variation, such vector-borne parasites are able to escape host immune responses, chronically infect and reinfect a broad host range and modulate the pathobiology of individual infections [4,9].

Typically, the within-host dynamics of these pathogens are characterized by the sequential emergence and replacement of different antigenic variants, creating an oscillating parasite load. Antigen switching is rooted in differential expression, one at a time, of members of a silent antigen gene archive, which in trypanosomes encodes variant surface glycoprotein (VSG), and in *P. falciparum* encodes *P. falciparum* erythrocyte membrane protein 1 (PfEMP1). The trypanosome archive has more than 1600 VSG genes [10] whose appearance follows a hierarchy, such that certain variants are activated preferentially over others, which is thought to have evolved as a core mechanism for optimizing the use of the archive [11]. In contrast to the huge family size of VSG in trypanosomes, PfEMP1 in *Plasmodium* is encoded by a family of roughly 60 *var* genes, involved in antigenic variation, virulence and adhesion to host receptors.

Across pathogens, there is evidence that hierarchy in antigen expression arises primarily via pathogen genetic factors governing gene switching, followed by immune selection [2,12–14]. Such factors include the chromosomal position of different antigen gene locus types [13], subtle differences in the flanks of the genes [2], and degree of identity between coding sequences [10,15]. Building on existing models of antigenic variation [16–22], there is a real opportunity now to develop a transdisciplinary framework for synthesizing a new understanding of the mechanistic links between archive architecture, pathogen population dynamics within hosts and transmission between host individuals.

Table 1. Bottom up determinants of antigen switch rates.

	genetic factors involved
repulsiveness	The probability of a gene to switch off may depend on genetic factors such as chromosomal location and transcription instability. For trypanosomes, early VSG inactivate at the rate of unrepaired expression site breakage, modified by the rate of recopying the same silent gene into the site, while later VSG inactivate by becoming converted by related silent genes. In *P. falciparum*, *var* genes have intrinsic inactivation frequencies, which appear to be related in part to complex forward and backward patterns in the switching lattice that links all of the repertoire [26]. There is also a position effect, such that *var* genes located in central chromosomal regions are expressed stably, possibly influenced by their promoter type, while those at subtelomeres switch off more frequently [27].
attractiveness	The intrinsic probability of a gene to be activated, independently of the gene being expressed, may depend on various factors, such as chromosomal location or number of repeats in flanking regions, as shown for *Trypanosoma brucei* [11]. In the simpler case of the single, circular bacterial chromosome, another factor that may contribute to the attractiveness of a particular gene is its spatial distance from the expression site: the further away, the lower its chances of being activated (e.g. pseudogenes in the bacterium *Anaplasma phagocytophilum* [28]).
mutual relatedness	The third component of the gravity model relates exclusively to mutual properties of the gene pair. Intragenic conversion of a segment from a silent gene or pseudogene into an expression site is a common theme in antigenic variation of a number of parasites and bacterial pathogens [29], and it seems to be highly favoured by the degree of coding sequence identity between the currently expressed and the new gene. A classical example is the formation and expression of mosaic VSG genes in the chronic stages of trypanosome infections [10]. This would make ϕ a decreasing function of mutual genetic distance, for example, a power-law, or a decaying exponential.

The recent availability of parasite genome data is revealing the structure and fine-scale architecture of antigen gene archives [10,23,24], thus calling for an evolutionary perspective to understand such complexity. In this article, we start this process, laying down a blue print for how these ideas might be linked together. We suggest that parasite genetic information, integrated with dynamic models of the within-host ecology and transmission, could be used to reveal the adaptive significance of archive structures and generate fresh insights about their function and deeper evolutionary ecology.

2. Material and methods

(a) From parasite genetics to antigen switch rates

Regardless of the specific genetic mechanisms involved, the event of switching from expression of one gene to another can depend on at least three types of factor: (i) properties of the gene that is being switched off; (ii) properties of the gene that is being activated; and (iii) shared properties between the two. Therefore, the rate at which gene j switches to gene i can be seen as a function of the individual contributions of these factors. Borrowing the notion of a gravity model [25], widely used in economic and traffic models, we propose to describe switch rates between two antigen genes as

$$s_{ji} = f(R_j, A_i, \phi(d_{ji})), \tag{2.1}$$

where R_j ('repulsiveness' of the current gene) describes the intrinsic propensity of a gene to switch away, A_i ('attractiveness' of the new gene) describes the intrinsic probability that a gene is activated and $\phi(d_{ji})$ is a function of the mutual relatedness between the two. In the context of antigenic variation, mutual relatedness d_{ji} can be coding sequence identity between the two genes, or regions of the two genes, or any other mutual feature with direct influence on their capacity to interact and recombine. Such recombination-driven switching has been observed for trypanosomes in the chronic stages of infection

when mosaic genes get expressed [10]. Typically, in a gravity model, the function f involves a product of the three components, R_j, A_i and ϕ. In table 1, we outline how these factors can depend on molecular mechanisms acting at the level of the parasite genome, and how they might vary from gene to gene and across pathogens.

To specify such a model, it is essential to quantify empirically features of antigen genes individually, and in higher-order groupings in a bottom-up fashion [26], and this remains a major challenge. While understanding such bottom-up mechanisms may enable genetic control and possibly intervention into the parasite switching process at high resolution, consideration of the macro-structure emerging in the archive from *all* switch rates is also important for understanding the role of the archive in parasite dynamics and fitness. Frank [19] has noted that selection on the archive might not operate at the level of individual switch rates, but rather on more collective structural properties, such as their magnitude ranges.

Recent empirical findings from antigen repertoires of different pathogens point towards a clustered genetic architecture composed of blocks of highly intra-connected variants and low between-block switch rates [11,23]. Based on this view, we propose a simplified description of the antigen archive using only three parameters: the number of genes in a single block (block size, η), the number of such blocks N_{blocks} and $\epsilon(<1)$, the average between-/within-block switching ratio (see the electronic supplementary material, S1 for details). Variation in any of these archive 'traits' results from fine-tuning of many entries s_{ij} of the switch matrix and has non-trivial implications for parasite within-host dynamics and transmission.

(b) From switch rates to within-host dynamics

The dynamics of each variant depend on the kinetic interactions between parasite cells and host immunity. The model we use is tailored to trypanosome antigenic variation [22], but has some parallels with other pathogen systems, such as *P. falciparum*. For example, the parasite population is subdivided between
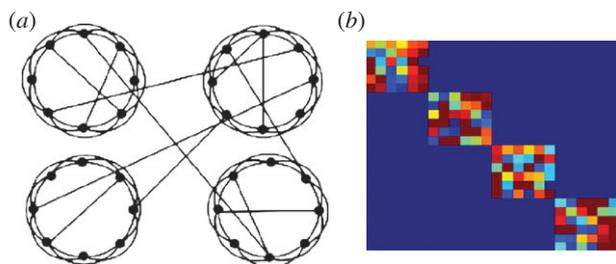
Figure 1. Illustration of an antigen gene block structure. (*a*) Antigen archive architecture. Switching within blocks of genes happens at a higher rate than switching between blocks. The magnitude of each switch rate (edge between nodes) can be formulated in terms of a gravity model, and hence possibly estimated directly from parasite genetics. (*b*) Antigen switch matrix. A typical switch matrix used in our models, $S_{N \times N} = s_{ij}$, is displayed as a heatmap, where red entries, within blocks, represent larger magnitude switch rates obtained through simulation, than entries between blocks, given by the blue range. See the electronic supplementary material, S1 for details on the precise construction of S.

N antigenically distinct variants and each variant has two life stages, in our case replicating (slender) and non-replicating (stumpy). We consider infection dynamics after initiation with a single variant. New variants are generated through antigen switching of existing parasites. Variant emergence is stochastic, governed by an underlying Markov process. As the infection proceeds, we track the probability that any variant has not yet been generated by time t. This probability crucially depends on the switch rates from other variants, assembled into a block structure matrix (figure 1).

When the generation probability of a new variant exceeds a certain threshold, that variant is expressed and starts to grow. Slender cells divide and are subject to antibody-mediated killing in a mass-action fashion. Furthermore, they can revert to stumpy form through density-dependent differentiation as the total population size increases. Stumpy cells instead do not divide and have a fixed lifespan; this enables the total parasite population to settle at the within-host carrying capacity (K) in the absence of immune control. Stumpy cells are killed by antibodies less efficiently than slender cells, with the consequence that, once antibodies arise, the non-replicating transmissive parasite forms persist longer in the host. Specific immunity receives equal stimulation from both cell types, with a characteristic delay. As a variant grows, the immune response grows in a saturating manner from 0 (assuming no prior exposure), up to a maximum of 1. Persistence of variant-specific immunity, even in the absence of antigen stimulation, prevents re-appearance of the same variant within the infection. For full model details refer to the electronic supplementary material, S1.

(c) Within-host dynamics and parasite fitness in one host

The net duration of host infectiousness, determining parasite local transmission success, is equal to the probability of the host to transmit the infection $\beta(t)$, weighted by the instantaneous host survival probability $\varphi(t)$, and integrated over the entire infection period. This index of parasite fitness in a single host (infection fitness) is given by

$$F = \int_0^\infty \beta(t)\varphi(t)\,\mathrm{d}t, \qquad (2.2)$$

where β and φ depend directly on parasite dynamics within the host. Quantifying this dependence empirically and theoretically

is crucial and lies at the heart of nested modelling [30]. The composition of archive blocks, and links between them, reflected in the switch matrix, can be key to the number, size, shape and separation of parasitaemia peaks throughout infection. These influences arise from the interplay between host and parasite parameters [22]. When the parasite expresses many blocks sequentially, several parasitaemia peaks develop in the form of block waves. For each block, peak parasite load and duration are determined by intrinsic growth rates of the constituent variants, kinetics and extent of differentiation to the non-dividing transmission stage, and variant-specific clearance mediated by host immunity. The lower the switch rates between blocks, the more decoupled the dynamics of each block and the greater the separation between subsequent peaks. All these characteristics influence parasite transmission probability to the vector, β, and host survival, φ. Evidence from vector-borne parasites [31–33] suggests that β might simply depend on a threshold number of transmissible parasites, needed to guarantee infection of the vector. In formulating host survival probability, φ, from within-host dynamics, we assume it depends only on host's natural mortality rate, and the cumulative total parasitaemia experienced over infection, multiplied by a *per capita* virulence factor (see the electronic supplementary material, S2–S3 for details and applications).

(d) From fitness in one host to fitness in the field

For parasites that can infect many hosts and alternate between hosts and vectors, overall fitness in the field is determined by these transmission events [34]. The basic reproduction number, R_0, in a heterogeneous host population becomes the sum of the contributions by all host types to the transmission cycle of the parasite:

$$R_0 = \frac{Z}{\gamma} \sum_{j \in \text{host types}} \frac{\alpha_j b_j^2 F_j}{H_j}, \qquad (2.3)$$

where Z is the abundance of vectors, H_j the abundance of each host type, α_j the probability a host of type j becomes infected from an infectious bite, γ vector mortality rate, b_j denotes the *per capita* vector bite rate on host type j and F_j is the average within-host parasite fitness in hosts of type j (see the electronic supplementary material, S4). If there is vertical transmission of the parasite, the expression for R_0 will have an additional term that explicitly sums the number of secondary cases generated via vertical transmission in the host population. Since the extent to which vertical transmission occurs in trypanosomes is unknown [35], we do not include it in this formulation. Equation (2.2) can be used to calculate F_j under the assumption that hosts of the same type are characterized by the same within-host parasite dynamics and epidemiological parameters. If $R_0 > 1$, the parasite can maintain itself after introduction into a totally susceptible host population. It has been argued that evolution of multi-host parasites depends on the interplay between constraints at the within- and between-host level [36]. A globally optimal parasite trait maximizes R_0 across a host range, while a local optimum maximizes fitness in a particular host type F_j. Naturally, divergent selective pressures on specific parasite traits get resolved depending on the relative densities of host types and their contributions to the transmission cycle of the pathogen. When certain host classes participate very little in the transmission cycle, the optimal global strategy of the parasite (e.g. antigen archive structure) will be closer to the local optima in the other classes.

(e) Quantifying host heterogeneity

The importance of host heterogeneity in parasite dynamics and evolution has received increasing attention in recent years, both
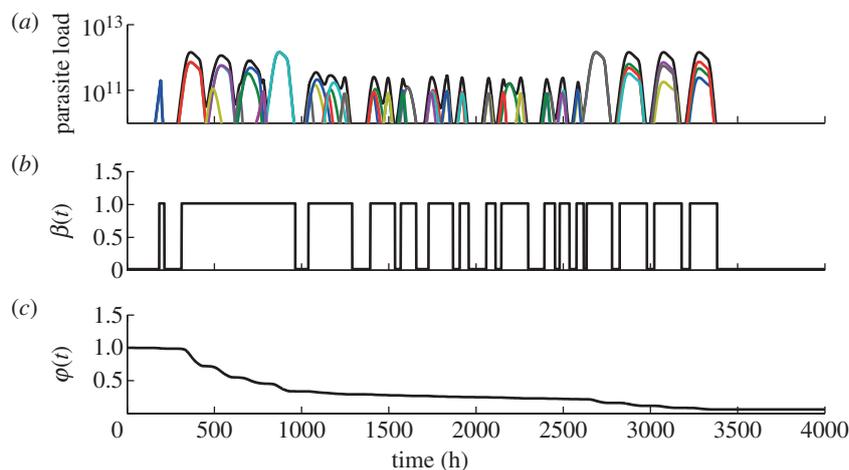
Figure 2. Antigenic variation dynamics integrated within a transmission framework. (*a*) Within-host parasite dynamics combined with (*b*) transmission potential, $\beta(t)$, and (*c*) host survival probability, $\varphi(t)$, yield parasite fitness from a single infection. The colours in the top graph depict individual antigen variants growing in the host. Here $F = \int \beta(t)\varphi(t)\,\mathrm{d}t = 757.24$. Parameter values as in the electronic supplementary material, tables S1–S2 with: $K_{\max} = K = C = 10^{12}$, $T = 4000$, $\eta = 3$, $N_{\text{blocks}} = 22$, $\epsilon = 0.001$, $\mu = 8.8 \times 10^{-3.5}$.

empirically and through theoretical models. Usually, the focus is on only one trait, or dimension of variability, although the actual diversity often spans multiple axes. Rather than considering changes in infection kinetics parameters between hosts in an isolated manner, neglecting possible correlations between them, it may be more appropriate and convenient to consider these characteristics as inter-dependent or directly linked to a host trait. Size is a key trait and important axis of variability between species, since it allometrically scales with many important physiological characteristics, such as metabolic rate, complexity, body temperature and lifespan [37]. Larger hosts are expected to have longer life expectancy, higher blood volume, smaller reproductive rate and lower population densities when compared with smaller host species.

Motivated by recent uses of allometric scaling to describe certain quantitative aspects of within-host and between-host infection dynamics [38–41], we extend our models, to study how host body mass (W) modulates the fitness effects of the parasite antigen archive. As an example, in calculating within-host parasite fitness $F(W)$, based on allometry, we assume that: (i) the within-host carrying capacity is proportional to host blood volume, thus to its mass W; (ii) natural host mortality scales as $W^{-1/4}$ [37]; (iii) the *per capita* increase in host mortality from one parasite cell scales also as $W^{-1/4}$ [40]; and (iv) other kinetics parameters, including immune response and parasite intrinsic growth rate are host size-invariant [38]. Despite these assumptions being inevitably simplistic, such an approach opens the way to new insight and interesting questions about the macro-forces governing the evolutionary dynamics of multi-host parasites and their immune evasion strategies.

3. Results

Applying the above modelling framework to the antigenic variation system of African trypanosomes, we explore how the infection characteristics mediated by the structure of the antigen archive affect parasite fitness (transmission potential) beyond the single host (figure 2). Trypanosome hosts often differ in many traits: their resistance to infection, the maximum parasite load they support (carrying capacity) or the pathogenesis they experience [42–44]. A natural question then is: how might such diversity of potential hosts impact on the evolution of parasite archive traits, e.g. block size and between-block connectivity? It is generally accepted

that parasites evolve to maximize their basic reproduction number [45]. This may lead to specialization in a single host species, using an archive that generates locally an optimal infection profile. We, therefore, investigate how the structure and differential expression of the archive can adapt to maximize infection fitness F in a given host that displays a particular set of parameters.

(a) Is there an optimal archive structure?
The within-host model reveals that when many variants grow at the same time, their individual numbers are kept low by density-dependent parasite differentiation, yielding suboptimal-specific immune responses and thus long parasite persistence. The size of an antigen block (η), determining the number of variants present in one parasitaemia wave, thus plays a pivotal role in mediating the regimes of rapid clearance and long persistence of a block, the latter leading to stationary infection as opposed to an oscillating one [22].

In order for the parasite to balance transmission potential with host mortality, there is an optimal antigen block size (figure 3*a*). Keeping all epidemiological parameters fixed, increasing η initially increases fitness, as the additional transmission obtained by longer parasite persistence outweighs the disadvantage caused by higher host mortality. However, further increases in block size lead to loss of variant-specific control and persistence of all variants, which can reduce host survival by a larger amount than it increases transmission probability, a trade-off that renders within-host fitness maximal at an intermediate value of η, the optimal block size η_{opt}. This effect is owing to the saturating nature of the transmission function, typical of vector-borne pathogens: as long as the number of parasites exceeds the transmission threshold, their exact number does not matter [31–33]. Thus, the excess parasite load above the transmission threshold, gained after a certain block size, acts only against the parasite.

In contrast to the single block size, a peak in parasite fitness as a function of the number of archive blocks is not found, there is no optimal N_{blocks}. Increasing the number of blocks increases within-host parasite fitness in a saturating manner, up to the point when further blocks of variants do not produce any additional gains in transmission (figure 3*b*). This maximal
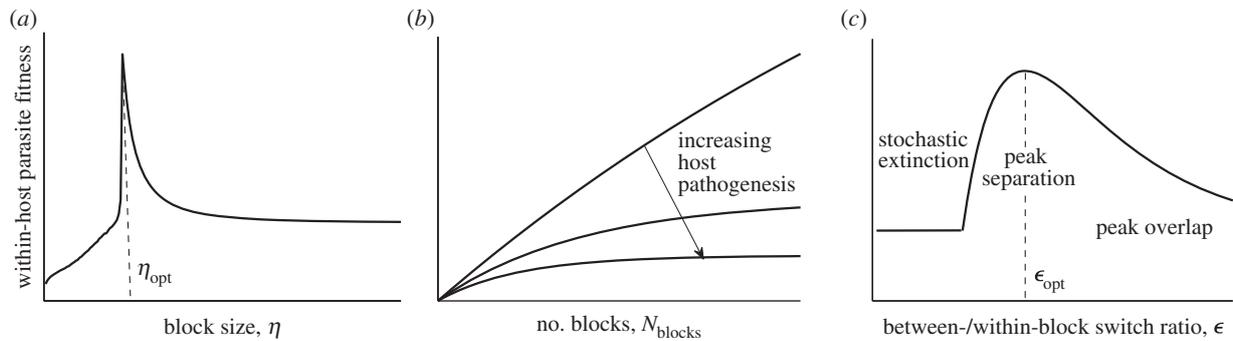
Figure 3. Within-host parasite fitness as a function of antigen archive traits. (*a*) An intermediate block size η_{opt} maximizes within-host parasite fitness, resolving the transmission-virulence trade-off. (*b*) Parasite fitness saturates with the number of archive blocks. (*c*) Within-host parasite fitness is maximal at an intermediate value of between-block switch rate.

fitness and the rate at which it is approached depend primarily on host natural and parasite-induced mortality. Intuitively, this is because the marginal effect of each new block of variants on parasite transmission declines as host survival probability declines. Archive modularity is most advantageous to the parasite early in infection, and more so if each individual block wave is relatively short (e.g. fast immune activation, high clearance rate, illustrated in the electronic supplementary material, S3). Later in infection, expression of additional blocks leads to smaller increases in fitness and is more advantageous if each block wave is longer. In general, once some level of archive modularity is reached, further expansion of the number of blocks does not contribute to within-host parasite fitness.

Between-block switch rate is responsible for the generation, temporal separation and asynchrony between blocks of parasite variants. The average ratio of between-block switch rates and within-block switch rates (ϵ) specifies the relative ease of accessing antigenic novelty within one group of genes versus across groups. If between-block switch rate is too low, demographic stochasticity combined with a limited reproduction potential in the ongoing block wave, prevents the next block from being generated [22]. If instead switching between blocks is very rapid, consecutive blocks appear in the same infection peak. Thus, there is an optimal between-/within-block switch ratio, where within-host parasite fitness is maximized at a balance between stochastic generation and temporal separation of variants (figure 3*c*).

(b) What factors shape the optimum?

A notable feature of η_{opt} is that for high virulence levels, small variations away from this optimum have a strong impact on fitness (figure 3*a*), indicating that the cost of expressing a block size above or below the optimum is very large. Importantly, small increases in block size from η_{opt} have a smaller effect on parasite fitness than small decreases, thus there may be marked pressure for block size not to be deficient. Interestingly, when the *per capita* host mortality induced by a single parasite is small, parasite fitness is considerably higher in a stationary infection scenario mediated by large antigen block sizes, than in an oscillatory infection (small block sizes), despite the larger overall parasite burden in the former. By contrast, when parasite-induced host mortality is high, infection fitness is higher if the block size is small, yielding an oscillatory infection (see the electronic supplementary material, S3).

Of several factors that might select for single block size, the strength of variant-specific versus density-dependent control within the host is most crucial. The more immune-competent the host, the larger the blocks it can control, and a parasite adapting to such a host breed or species must increase the size of its blocks. For example, in a large host where the ratio between the carrying capacity and the antigen detection threshold K/C is 1, the optimal block size seems to be 43 variants, but if the antigen detection threshold decreases relatively by 50 per cent, increasing specific immune control in that host, the optimal block size for the antigen archive increases to 67 (e.g. electronic supplementary material, figure S7*a*). Interestingly, host–parasite coevolution could drive towards compensatory increases in efficiency of immunity (lower threshold for detection of antigen) and in antigen block size. For a parasite such as the trypanosome, facing a broad host-range and therefore heterogeneity in immune-competence, a larger size for antigen gene blocks may be selected predominantly by the more competent hosts. If such an optimal archive trait can be acquired at no physiological cost, the parasite should evolve towards η_{opt}, when adaptation to these hosts prevails. However, if additional synchronization in variant expression comes with intrinsic costs, the parasite may not be able to evolve towards the optimal block size. In that scenario, immune-competent hosts may constrain single block size evolution and confer sub-optimal fitness to the parasite. Moreover, a pathogen that has evolved one block size could be hypervirulent in a second host species that is less immune-competent, such hypervirulence being driven by over-riding selection in the dominant host species.

When the size of a single block is moderate, it is exactly the number of such blocks that determines the distinction between an acute (single parasitaemia peak) and chronic infection (multiple peaks). The relative gain in parasite fitness owing to shifting from a short to a long infection is indeed larger in tolerant than in sensitive hosts. Furthermore, the number of blocks N_{blocks} interacts also with the differentiation rate, modulated by the within-host carrying capacity K. The distinction in parasite fitness between a severe acute ($N_{blocks} = 1$, large K) and a mild chronic ($N_{blocks} > 1$, small K) infection is dependent on the vector transmission threshold (see the electronic supplementary material, S3). As the transmission threshold increases, requiring higher parasite loads to infect the vector, parasite fitness tends to favour the severe acute infection versus the mild chronic one. It is thus conceivable that the transmission threshold,

an intrinsic property of the vector, can also play a role in the coevolution of these two parasite traits: the number of archive blocks and slender-to-stumpy differentiation rate.

Regarding the rate of between-block switching, we observe that its optimal value is reduced if the replication potential of a single block increases, for example, if the delay in immune response activation and/or within-host carrying capacity increases. A larger delay implies longer duration of the first block wave before clearance, whereas higher carrying capacity yields a higher peak. Since stochastic variant generation depends directly on the number of replicating cells in the ongoing block wave, new blocks emerge more easily when the absolute growth limit set by the within-host carrying capacity is higher, and the parasite can then afford to switch more slowly (e.g. electronic supplementary material, figure S7b).

If the carrying capacity is fixed, it is the partitioning of the parasite population between reproductive and transmissive stages within a block that matters. The transmission-reproduction ratio within a block affects not only current transmission, but also future transmission by indirectly altering the rate of antigen turnover for future blocks, as has been noted earlier [46]. To optimize the progression of antigenic variation within a host and ensure onward transmission, the parasite must adopt intermediate rates of switching between blocks and an intermediate ratio between reproductive and transmissive stages (see the electronic supplementary material, S3).

Between-/within-block switch ratio (ϵ) impacts also on the parameter regime for which a large block size is favoured over a small one. For example, as ϵ increases and we move from no subsequent block generation, to block separation and to block overlap, the virulence range favouring stationary infection, as opposed to an oscillatory one (i.e. large block size), first decreases and then increases, tending to a constant. This suggests that the fitness benefits from expressing a given block size, η, are maximal for an intermediate value of between-block switch rate, and highlights the interdependence between different antigen archive parameters.

(c) When factors covary: the influence of host body size
In the allometrically scaled model, where host parameters such as tolerance to infection and parasite carrying capacity vary together, we find that as the size of the host increases, within-host parasite fitness first increases, owing to the lower pathogenesis rates and better host survival. However, if the host size increases further, the corresponding increase in carrying capacity and antigen turnover rate reduces within-host fitness as the archive is expressed too quickly, shortening the overall infection. This reduction in the transmission window of the parasite is apparently larger than the benefit of lower pathogenesis in larger hosts, although the exact magnitude of this effect is likely to depend on the precise scaling involved. However, for a parasite with a broad range of hosts of different sizes, there seems to be an optimal intermediate size of host, in which to maximally exploit a given archive (see the electronic supplementary material, S4).

Our results further show that the sensitivity of within-host fitness, F, to changes in the number of blocks, N_{blocks} and the ratio of between- and within-block switch rates (ϵ) increases with host body size, suggesting that the evolutionary dynamics

of these parasite antigen archive parameters should exhibit tight coupling to this host trait. Under the assumption that the basic kinetics of processes in immune systems are scale-invariant (e.g. the antigen detection threshold, relative to within-host carrying capacity, does not change with host size), the sensitivity of F to the size of a single block (η) in the archive should remain generally host size independent.

We expect at least two ways in which the archive may evolve in adapting to larger size hosts: more blocks of variants and more separation between them. A plausible explanation for the large size and modular structure of the extant trypanosome antigen archive may partly be the large selection pressure for these archive traits, coming typically from large hosts, dominant in the transmission cycle of this parasite. Indeed, the exceptionally broad host range might contribute to the fact that the number of antigen genes in the trypanosome archive vastly exceeds that of other parasites, particularly when the added effect of combinatorial creation of mosaic expressed genes is considered.

4. Discussion
(a) Linking scales in antigenic variation
Here, we used a nested modelling framework to illustrate, for trypanosomes, the role that the antigen archive structure plays in determining parasite success in a single host and how this could affect between-host dynamics. A trade-off between transmission and virulence naturally arises in the context of antigenic variation and intermediate values for the size of antigen blocks and between-block switch rates appear optimal. Theoretical models have long postulated that parasites might evolve intermediate levels of virulence [47–49], differentiation between parasite reproductive and transmission stages [50,51] or mutation rates [52]. Extending these ideas to a novel context, the antigen archive structure as an important pathogen trait under natural selection, we explored alternative strategies for fitness optimization in antigenically varying pathogens.

Although our study highlights the qualitative influence of antigen archive parameters on parasite fitness and its links across biological scales, a more quantitative approach is needed in the future. For example, establishing the precise functional dependence of parasite transmission potential on archive structure (η, N_{blocks}, ϵ) as a whole, taking into account possible covariance and interactions between its components remains a challenge. Our baseline formulation assumed that cross-reactive immunity does not contribute significantly. An important avenue for further exploration is the interaction between cross-immunity and immune exhaustion in determining the success of a given antigen archive structure, as it has been recently recognized that these phenomena might play a key role in persistent infections [53].

Antigenically varying parasites are subject to a diverse range of selection pressures, one of which is the immune memory of their hosts at the population level. Previous studies, focusing on repertoire composition and the antigenic features of particular variants, have explained the maintenance of parasite strains with non-overlapping antigen repertoires as a result of such pressures [54]. More recently, embedding explicitly the within-host dynamics in between-host transmission and considering variation in dominance among variant surface antigens, the complex multi-layered

Table 2. Top-down influences on the antigen archive structure.

	parameter change	F	antigen archive evolution
host	antigen detection threshold $\downarrow$	$\downarrow$	$\eta \uparrow$
	immune response delay $\uparrow$	$\uparrow$	$\epsilon \downarrow$
	parasite clearance rate $\uparrow$	$\downarrow$	$N_{blocks} \uparrow$, $\eta \uparrow$
	tolerance $\uparrow$	$\uparrow$	$N_{blocks} \uparrow$, $\eta \uparrow$
	within host carrying capacity $\uparrow$	$\uparrow$	$\eta \uparrow$, $\epsilon \downarrow$
vector	transmission threshold $\downarrow$	$\uparrow$	$\eta \downarrow$, $N_{blocks} \uparrow$

population dynamics created by reinfection, population turn-over and antigenic diversity has also been addressed with individual-based models for malaria [55,56]. Similar mechanisms could play a role also in trypanosomes, although the size of their genomic antigen repertoire and ability to generate novel VSG over single infections greatly surpasses that of *P. falciparum*, reducing the pressure for non-overlapping structure at the strain level. Differently from malaria parasites, in trypanosomes reinfection and its epidemiology have not yet been quantified in detail empirically.

Although host reinfection is not currently considered by our models, our primary focus being the structure of the antigen archive, more positive effects of archive expansion and modularity are expected to feedback if the parasite reinfects partially immune hosts. This is true for *Plasmodium* [26] but could also apply to trypanosomes. Since larger size and generally more tolerant hosts are likely to have been exposed to a larger proportion of the archive over any single infection, their reinfection would be facilitated by expression of new antigens. Thus, hosts with greater immune memory could select for more antigenic variability of the pathogen, either by selecting for additional new blocks in the archive, or for smaller blocks that would limit exposure diversity. Previous research has linked the divergent selective pressures from hosts of different immune memory to some pathogen variability in age-class bias, reproductive rate and antigenic variation [57,58]. By analogy, the divergent selective forces from hosts of different size could be related to differences between trypanosome strains in host-bias and antigen archive structure or composition, an area awaiting future studies.

Regardless of such limitations, our models point to various top-down factors that can act as evolutionary drivers for different structural aspects of the antigen archive (table 2), reinforcing the importance of host heterogeneity in parasite evolutionary dynamics [59,60]. The magnitudes of these selection pressures will inevitably depend on host demography such as relative abundances between host types (see the electronic supplementary material, S4 for an example), host susceptibilities and ecological factors such as contact rates with the vector, which have been shown to affect the evolution of multi-host pathogens [36], but yet remain a challenge empirically. Naturally, divergent selection

on components of antigen repertoires would have to be resolved depending on such differential contributions to the transmission cycle of the pathogen.

(b) Rethinking pathogen evolution

Despite the emphasis on trypanosomes, our study suggests that there are general larger-scale consequences of parasite genetic architecture that influence within-host and between-host population dynamics. Indeed, the selection forces that shape the parasite genome are likely to be determined primarily by these larger-scale processes. The optimal way in which parasites should structure their antigen archive is dependent on between-host dynamics and the ecological and epidemiological factors driving them. The results of these processes lie encrypted within the contemporary parasite genomes that we study today. In this post-genomic era, we are able to deduce details of parasite genome architecture that govern the generation of antigenic novelty within and between hosts, and the factors possibly constraining this process at the genetic level. It might be that certain antigen archive traits are too hard to develop genetically, and that alternative life-history traits, such as reproductive restraint [61], or sensitivity to differentiation [62], could evolve instead in order to yield equivalent fitness benefits to the parasite. The difficulty is that we do not know what life-history alternatives are available to pathogens, and unless those options are known, prediction of their adaptive value is hard.

Increasing knowledge of parasite genetic details will enable integration of population dynamics and genetic processes, and the synthesis should stimulate new approaches to how we think about within- and between-host pathogen population dynamics. As these higher-level processes exert profound selective pressures on the structure and use of antigen archives, uncovering the inter-dependencies across different levels of biological organization becomes key to understanding how parasites have evolved to be the way they are.

References

1. Donelson J. 1995 Mechanisms of antigenic variation in *Borrelia Hermsii* and African trypanosomes. *J. Biol. Chem.* **270**, 7783–7786.

2. Barbour AG, Dai Q, Restrepo BI, Stoenner HG, Frank SA. 2006 Pathogen escape from host immunity by a genome program for antigenic variation. *Proc. Natl Acad. Sci. USA* **103**, 18 290–18 295. (doi:10.1073/pnas.0605302103)

3. Dai Q, Restrepo BI, Porcella SF, Raffel SJ, Schwan TG, Barbour AG. 2006 Antigenic variation by *Borrelia hermsii* occurs through recombination between extragenic repetitive elements on linear plasmids. *Mol. Microbiol.* **60**, 1329–1343. (doi:10.1111/j.1365-2958.2006.05177.x)

4. Deitsch KW, Lukehart SA, Stringer JR. 2009 Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat. Rev. Microbiol.* **7**, 493–503. (doi:10.1038/nrmicro2145)

5. Vink C, Rudenko G, Seifert HS. 2012 Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiol. Rev.* **36**, 917–948. (doi:10.1111/j.1574-6976.2011.00321.x)

6. Barbour AG, Restrepo BI. 2000 Antigenic variation in vector-borne pathogens. *Emerg. Infect. Dis.* **6**, 449–457. (doi:10.3201/eid0605.000502)

7. Frank S. 2002 *Immunology and evolution of infectious disease*. Princeton, NJ: Princeton University Press.

8. Lipsitch M, O'Hagan J. 2007 Patterns of antigenic diversity and the mechanisms that maintain them. *J. R. Soc. Interface* **4**, 787–802. (doi:10.1098/rsif.2007.0229)

9. Turner M. 1999 Antigenic variation in *Trypanosoma brucei* infections: an holistic view. *J. Cell Sci.* **112**, 3187–3192.

10. Marcello L, Barry J, 2007 Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favoured by archive substructure. *Genome Res.* **17**, 1344–1352. (doi:10.1101/gr.6421207)

11. Morrison L, Marcello L, McCulloch R. 2009 Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. *Cell. Microbiol.* **11**, 1724–1734. (doi:10.1111/j.1462-5822.2009.01383.x)

12. Barry J, McCulloch R. 2001 Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv. Parasitol.* **49**, 1–70. (doi:10.1016/S0065-308X(01)49037-3)

13. Morrison L, Majiwa P, Read A, Barry J. 2005 Probabilistic order in antigenic variation of *Trypanosoma brucei*. *Int. J. Parasitol.* **35**, 961–972. (doi:10.1016/j.ijpara.2005.05.004)

14. Kyes S, Kraemer S, Smith J. 2007 Antigenic variation in *Plasmodium falciparum*: gene organization and regulation of the *var* multigene family. *Eukaryot. Cell* **6**, 1511–1520. (doi:10.1128/EC.00173-07)

15. Barbet AF, Kamper S. 1993 The importance of mosaic genes to trypanosome survival. *Parasitol. Today* **9**, 63–66. (doi:10.1016/0169-4758(93)90039-I)

16. Agur Z, Abiri D, Van der Ploeg LH. 1989 Ordered appearance of antigenic variants of African trypanosomes explained in a mathematical model based on a stochastic switch process and immune-selection against putative switch intermediates. *Proc. Natl Acad. Sci. USA* **86**, 9626–9630. (doi:10.1073/pnas.86.23.9626)

17. Nowak M, May R, Anderson R. 1990 The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS* **4**, 1095–1103. (doi:10.1097/00002030-199011000-00007)

18. Antia R, Nowak MA, Anderson R. 1996 Antigenic variation and the within-host dynamics of parasites. *Proc. Natl Acad. Sci. USA* **93**, 985–989. (doi:10.1073/pnas.93.3.985)

19. Frank S. 1999 A model for sequential dominance in African trypanosome infections. *Proc. R. Soc. Lond. B* **266**, 1397–1401. (doi:10.1098/rspb.1999.0793)

20. Recker M, Nee S, Bull P, Kinyanjui S, Marsh K, Newbold C, Gupta S. 2004 Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. *Nature* **429**, 555–558. (doi:10.1038/nature02486)

21. Lythgoe K, Morrism LJ, Read A, Barry J. 2007 Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. *Proc. Natl Acad. Sci USA* **104**, 8095–8100. (doi:10.1073/pnas.0606206104)

22. Gjini E, Haydon D, Barry J, Cobbold C. 2010 Critical interplay between parasite differentiation, host immunity, and antigenic variation in trypanosome infections. *Am. Nat.* **176**, 424–439. (doi:10.1086/656276)

23. Bull P, Buckee C, Kyes S, Kortok M, Thathy V, Guyah B, Stoute J, Newbold C, Marsh K. 2008 *Plasmodium falciparum* antigenic variation. mapping mosaic *var* gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol. Microbiol.* **68**, 1519–1534. (doi:10.1111/j.1365-2958.2008.06248.x)

24. Jackson A *et al*. 2010 The genome sequence of *Trypanosoma brucei* gambiense, causative agent of chronic human African trypanosomiasis. *PLoS Negl. Trop. Dis.* **4**, e658. (doi:10.1371/journal.pntd.0000658)

25. Sen A, Smith T. 1995 *Gravity models of spatial interaction behaviour*. Berlin, Germany: Springer.

26. Recker M, Buckee C, Serazin A, Kyes S, Pinches R, Christodolou Z, Springer A, Gupta S, Newbold C. 2011 Antigenic variation in *Plasmodium falciparum* malaria involves a highly structured switching pattern. *PLOS Pathog.* **7**, e1001306. (doi:10.1371/journal.ppat.1001306)

27. Frank M, Dzikowski R, Amulic B, Deitsch K. 2007 Variable switching rates of malaria virulence genes are associated with chromosomal position. *Mol. Microbiol.* **64**, 1486–1498. (doi:10.1111/j.1365-2958.2007.05736.x)

28. Foley J, Nieto J, Barbet A, Foley P. 2009 Antigen diversity in the parasitic bacterium *Anaplasma phagocytophilum* arises from selectively-represented, spatially clustered functional pseudogenes. *PLoS ONE* **4**, e8265. (doi:10.1371/journal.pone.0008265)

29. Palmer G, Brayton K. 2007 Gene conversion is a convergent strategy for pathogen antigenic variation. *Trends Parasitol.* **23**, 408–413. (doi:10.1016/j.pt.2007.07.008)

30. Mideo N, Alizon S, Day T. 2008 Linking within- and between-host dynamics in the evolutionary epidemiology of infectious disease. *Trends Ecol. Evol.* **23**, 511–517. (doi:10.1016/j.tree.2008.05.009)

31. Maudlin I, Welburn S. 1989 A single trypanosome is sufficient to infect a tsetse fly. *Ann. Trop. Med. Parasitol.* **83**, 431–433.

32. Van den Bossche P, Ky-Zerbo, A, Brandt J, Marcotty, T, Geerts S, De Deken R. 2005 Transmissibility of *Trypanosoma brucei* during its development in cattle. *Trop. Med. Int. Health* **10**, 833–839. (doi:10.1111/j.1365-3156.2005.01467.x)

33. Paul R, Bonnet S, Boudin C, Tchuinkam T, Robert V. 2007 Aggregation in malaria parasites places limits on mosquito infection rates. *Infect. Genet. Evol.* **7**, 577–586. (doi:10.1016/j.meegid.2007.04.004)

34. Diekmann O, Heesterbeek J, Metz J. 1990 On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382. (doi:10.1007/BF00178324)

35. Lindner AK, Priotto G. 2010 The unknown risk of vertical transmission in sleeping sickness: a literature review. *PLoS Negl. Trop. Dis.* **4**, e783. (doi:10.1371/journal.pntd.0000783)

36. Gandon S. 2004 Evolution of multi-host parasites. *Evolution* **58**, 455–469.

37. Peters R. 1983 *The ecological implications of body size. Cambridge studies in ecology*. New York, NY: Cambridge University Press.

38. Perelson A, Wiegel F. 2004 Some scaling principles for the immune system. *Immunol. Cell Biol.* **82**, 127–131. (doi:10.1046/j.0818-9641.2004.01229.x)

39. Dobson A. 2004 Population dynamics of pathogens with multiple host species. *Am. Nat.* **164**, S64–S78. (doi:10.1086/424681)

40. Cable J, Enquist B, Moses M. 2007 The allometry of host-pathogen interactions. *PLoS ONE* **2**, e1130. (doi:10.1371/journal.pone.0001130)

41. Bolzoni L, De Leo G, Gatto M, Dobson A. 2008 Body-size scaling in an SEI model of wildlife diseases. *Theor. Popul. Biol.* **73**, 374–382. (doi:10.1016/j.tpb.2007.12.003)

42. Morrison W, Murray M. 1985 The role of humoral immune responses in determining susceptibility of A/J and C57BL/6 mice to infection with *Trypanosoma congolense*. *Parasite Immunol.* **7**, 63–79. (doi:10.1111/j.1365-3024.1985.tb00479.x)

43. Barry J. 1986 Antigenic variation during *Trypanosoma vivax* infections of different host species. *Parasitology* **92**, 51–65. (doi:10.1017/S0031182000063447)

44. Taylor K. 1998 Immune responses of cattle to African trypanosomes: protective or pathogenic? *Int. J. Parasitol.* **28**, 219–240. (doi:10.1016/S0020-7519(97)00154-9)

45. May RM, Anderson R. 1983 Epidemiology and genetics in coevolution of parasites and hosts. *Proc. R. Soc. Lond. B* **219**, 281–313. (doi:10.1098/rspb.1983.0075)

46. MacGregor P, Savill N, Hall D, Matthews K. 2011 Transmission stages dominate trypanosome within-host dynamics during chronic infections. *Cell Host Microbe* **9**, 310–318. (doi:10.1016/j.chom.2011.03.013)

47. Levin S, Pimental D. 1981 Selection of intermediate rates of increase in parasite–host systems. *Am. Nat.* **117**, 308–315. (doi:10.1086/283708)

48. Anderson R, May R. 1982 Coevolution of hosts and parasites. *Parasitology* **85**, 411–426. (doi:10.1017/S0031182000055360)

49. Ewald PW. 1983 Host–parasite relations, vectors and the evolution of disease severity. *Annu. Rev. Ecol. Evol. Syst.* **14**, 465–485. (doi:10.1146/annurev.es.14.110183.002341)

50. Sasaki A, Iwasa Y. 1991 Optimal growth schedule of pathogens within a host: switching between lytic and latent cycles. *J. Theor. Popul. Biol.* **39**, 201–239. (doi:10.1016/0040-5809(91)90036-F)

51. Alizon S, van Baalen M. 2008 Transmission-virulence trade-offs in vector-borne diseases. *Theor. Pop. Biol.* **74**, 6–15. (doi:10.1016/j.tpb.2008.04.003)

52. Kamp C, Wilke C, Adami C, Bornholdt S. 2002 Viral evolution under the pressure of an adaptive immune system: optimal mutation rates for viral escape. *Complexity* **8**, 28–33. (doi:10.1002/cplx.10067)

53. Johnson P, Kochin B, Ahmed R, Antia R. 2012 How do antigenically varying pathogens avoid cross-reactive responses to invariant antigens? *Proc. R. Soc. B* **279**, 2777–2785. (doi:10.1098/rspb.2012.0005)

54. Gupta S, Maiden M, Feavers I, Nee S, May Rea. 1996 The maintenance of strain structure in populations of recombining infectious agents. *Nat. Med.* **2**, 437–442. (doi:10.1038/nm0496-437)

55. McKenzie F, Bossert W. 2005 An integrated model of *Plasmodium falciparum* dynamics. *J. Theor. Biol.* **232**, 411–426. (doi:10.1016/j.jtbi.2004.08.021)

56. Severins M, Klinkenberg D, Heesterbeek H. 2012 How selection forces dictate the variant surface antigens used by malaria parasites. *J. R. Soc. Interface* **9**, 246–260. (doi:10.1098/rsif.2011.0239)

57. Frank SA, Bush R. 2007 Barriers to antigenic escape by pathogens: trade-off between reproductive rate and antigenic mutability. *BMC Evol. Biol.* **7**, 229 (doi:10.1186/1471-2148-7-229)

58. Van Noort SP, Nunes MC, Weedall GD, Hviid L, Gomes MGM. 2010 Immune selection and within-host competition can structure the repertoire of variant surface antigens in *Plasmodium falciparum*: a mathematical model. *PLoS ONE* **5**, e9778. (doi:10.1371/journal.pone.0009778)

59. Regoes RR, Nowak MA, Bonhoeffer S. 2000 Evolution of virulence in a heterogeneous host population. *Evolution* **54**, 64–71. (doi:10.1111/j.0014-3820.2000.tb00008.x)

60. Ganusov V, Bergstrom CT, Antia R. 2002 Within-host population dynamics of microparasites in a heterogeneous host population. *Evolution* **56**, 213–223. (doi:10.1111/j.0014-3820.2002.tb01332.x)

61. Reece S, Ali E, Schneider P, Babiker H. 2010 Stress, drugs and the evolution of reproductive restraint in malaria parasites. *Proc. R. Soc. B* **277**, 3123–3129. (doi:10.1098/rspb.2010.0564)

62. Matthews K. 2011 Controlling and coordinating development in vector-transmitted parasites. *Science* **331**, 1149–1153. (doi:10.1126/science.1198077)